

Supplemental Figure Legends

Figure S1. Representative RNA quality (TapeStation) of mock RNA pool #1 as well as human stool, oral and vaginal microbiome samples used in this study.

A) Electropherograms for total stool RNA extracted with PowerMicrobiome and RNA pool # 1 following sample analysis on an RNA screentape. B) Electropherograms for total RNA extracted from OMNIgene•ORAL (OMR-120 –tongue microbiome) and OMNIgene•VAGINAL (OMR-130) collected on high sensitivity RNA screentape.

Figure S2. Detailed comparison of informatics tools for classifying rRNA reads from metatranscriptomes.

Bar plots showing fractions classified as either rRNA or 'clean' (mRNA) by the various tools (bow: Bowtie 2; duk: BBDuk; sea: Seal; smr: SortMeRNA; spl: bbsplit) searched against four different databases (ar: SILVA SSU/LSU NR99; art: ar + tRNA RFAM clans; ds: SortMeRNA (4.3) default database; ss: SortMeRNA (4.3) sensitive database). The line plots represent the rRNA content of the sample. B) Read depth for the samples sequenced in this study.

Figure S3. Detailed workflow and strategy for human microbiome rRNA depletion probe design.

A) Detailed flowchart of the iterative probe design process used to generate HMv1 and HMv2 custom rRNA depletion probe sets. The steps highlighted in blue are described in more detail in B) and C). B) Following depletion of donor samples with DP1 and alignment of fastq files to the SilvaDB, regions of rRNAs that have over 500x coverage are collected. A representative donor sample is visualized in IGV (upper panel) showing both undepleted and DP1 depleted sequencing reads with alignment of the DP1 probes across the bottom of the panel. Regions where DP1 probes do not align corresponds to high rRNA read coverage (dotted lines). The rRNA regions with >500x coverage were collected and their length was plotted vs frequency observed (two examples shown in lower panels). Most of these regions are <200 nt in length. C) A heatmap showing the top 50 most abundant regions from each donor that were pooled and pairwise alignment was performed. Red indicated regions with high percent alignment and blue represents low percent alignment. Gray represent zero percent alignment. If two regions aligned >80% only one region was chosen to reduce redundancy. Therefore, red areas of the heatmap resulted in fewer probes designed, while blue or gray areas required more probes to be generated.

Figure S4. Functional profiles of adult fecal metatranscriptomes.

A) RefSeq Functions. B) SEED Functions.

Figure S5. Comparing the rRNA depletion efficiency of the original Ribo-Zero Gold (Epidemiology) kits or Ribo-Zero Plus with DP1 or HMv1 probe sets from ATCC mock community samples.

A) Depletion from ATCC skin samples is more efficient when using the Ribo-Zero Gold (Epidemiology) kit (RZBact) than Ribo-Zero Plus (RZ + DP1). However, substituting DP1 with the HMv1 probe set results in efficient rRNA depletion (HMv1). B) Depletion of ATCC gut samples is effective using only DP1 and demonstrates that Ribo-Zero Plus has better reproducibility than the traditional hybrid capture method which is prone to contamination caused by inadvertent carry over of depletion beads during the protocol.

Figure S6. RT-qPCR based assessment of vaginal microbiome *Lactobacillus* content and additional oral rRNA depletion data.

A) RT-qPCR quantification of *Lactobacillus*-dominated and non-*Lactobacillus*-dominated samples and controls. B) rRNA depletion of additional tongue microbiome samples using the HMv1 probes.

Figure S7. rRNA depletion efficiency in stool samples of donors between 4 and 33 months old.

A) Percentage of reads mapping to eukaryotic and prokaryotic rRNAs vs coding sequences in infant and child stool samples depleted with Ribo-Zero Plus with HMv1 probes. Replicate rRNA depletion reactions were performed for each sample. B) Percentage of reads mapping to eukaryotic and prokaryotic rRNAs vs. other sequences in four infant stool samples depleted with Ribo-Zero Plus with HMv1 or HMv2 compared to undepleted samples. Replicate rRNA depletion reactions were performed for each condition. Boxed bar plot on top of panel B represents remaining bacterial rRNA (LSU & SSU) with a dashed line at the 50% mark.

Figure S8. Comparison of rRNA depletion efficiency of adult stool RNA samples using the HMv1 and HMv2 custom probe sets

Percentage of reads mapping to eukaryotic and prokaryotic rRNAs vs. other sequences in adult stool samples depleted with Ribo-Zero Plus with HMv1 or HMv2. the same 10 healthy adult samples were used as in Figure 3.

Figure S9. Comparison of number of features detected for infant samples depleted with different methods.

Number of reads, or taxonomic or functional features detected for undepleted, HMv1-depleted, or HMv2-depleted infant samples.

Figure S10. Infant versus adult metatranscriptomic profiles.

A) Adult and infant RefSeq taxa. B) Adult and infant CAZyme profiles.

Figure S11. Infant versus adult metatranscriptomic profiles.

Effects plot showing functional categories that exhibit differential expression between adult and infant samples.

Figure S12. Volcano plots of adult and infant taxa or function.

Taxa (A, C) or functional groups (B, D) that are significantly different in the indicated comparison are colored in red, non-significant differences are in black. A) Infant taxa from samples <6 months vs. >20 months, B) Infant RefSeq functions from samples <6 months vs. >20 months, C) Infant vs. adult taxa, D) Infant vs. adult RefSeq functions.

Figure S13. ERCC spike-in depletion and mapping data.

A) rRNA depletion for indicated ERCC spike-in samples. B) Proportion of reads aligning to all ERCC spike-ins in indicated samples. C) For the reads aligning to the ERCC spike-ins, the proportion of reads aligning to each ERCC spike-in for indicated samples.

Figure S14. Pairwise between-group Spearman correlations

Pairwise average Spearman correlations of ERCC spike-in data between the different groups tested.

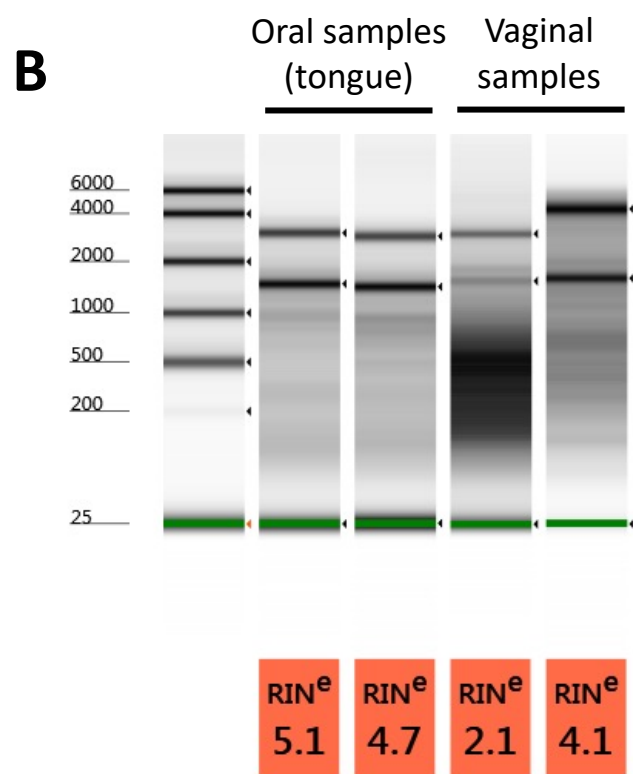
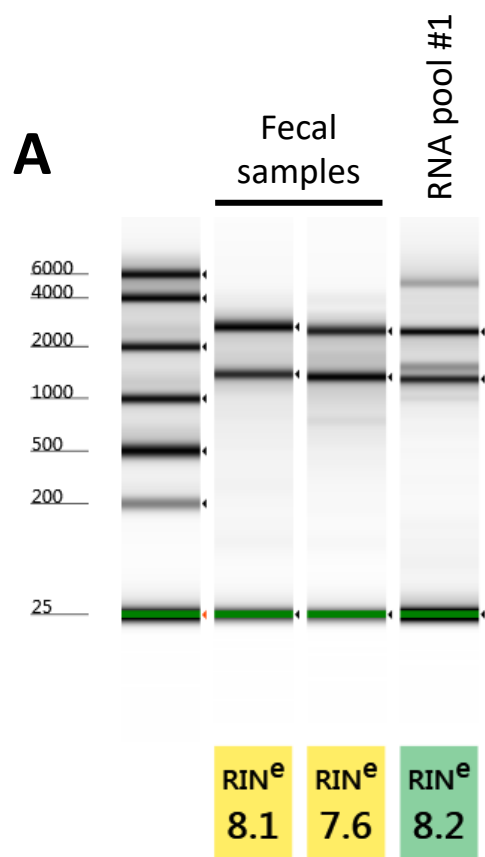
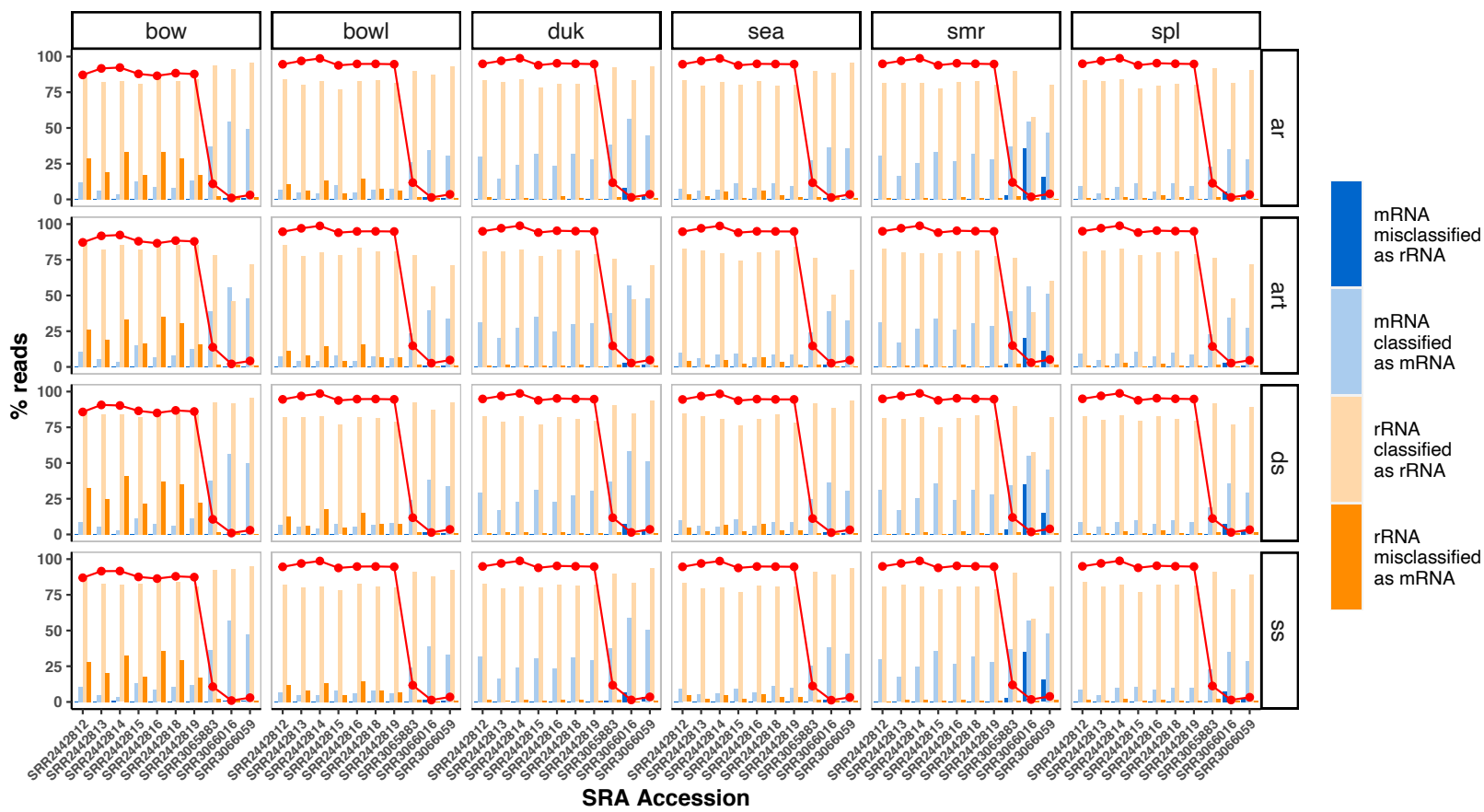


Figure S1

A



B

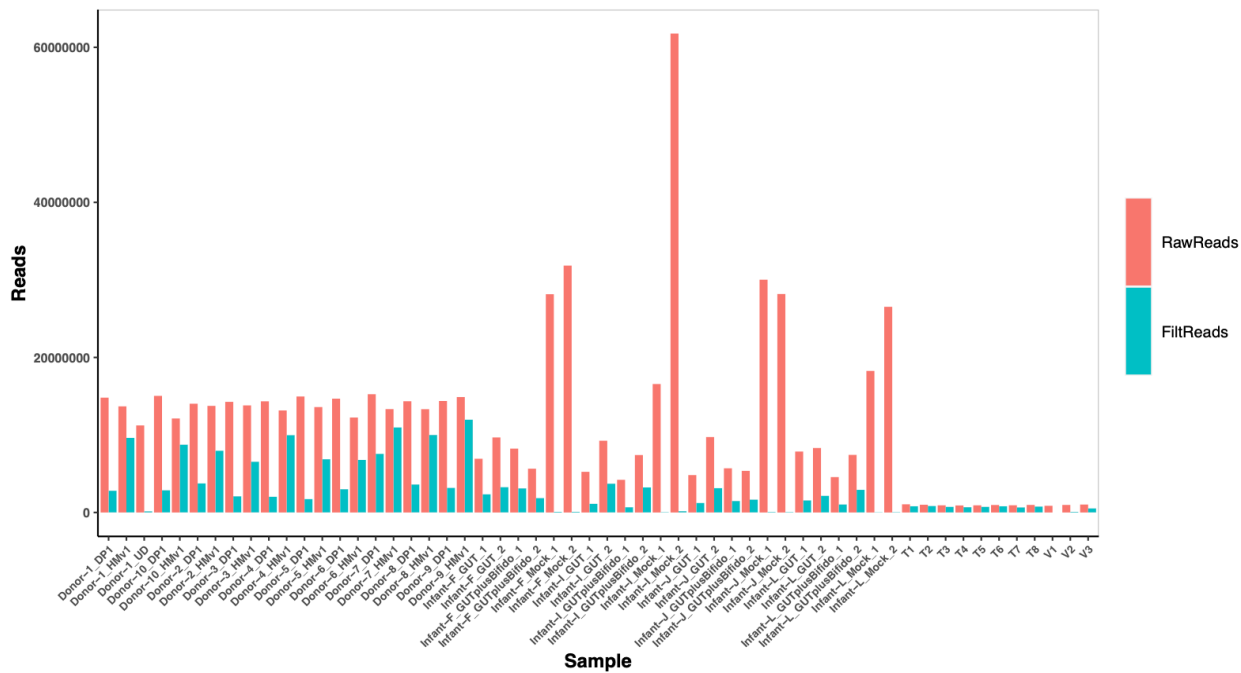


Figure S2

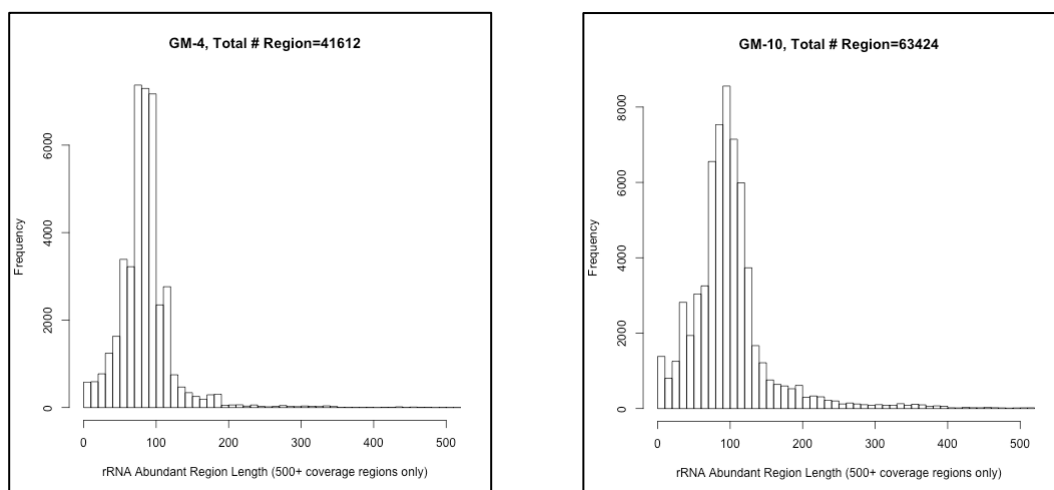
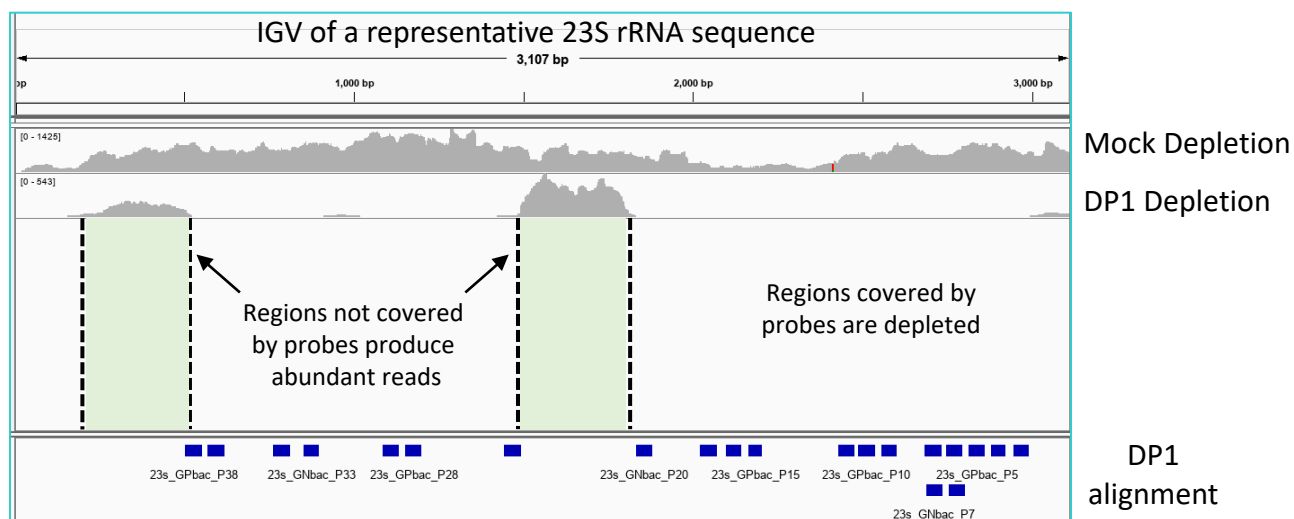
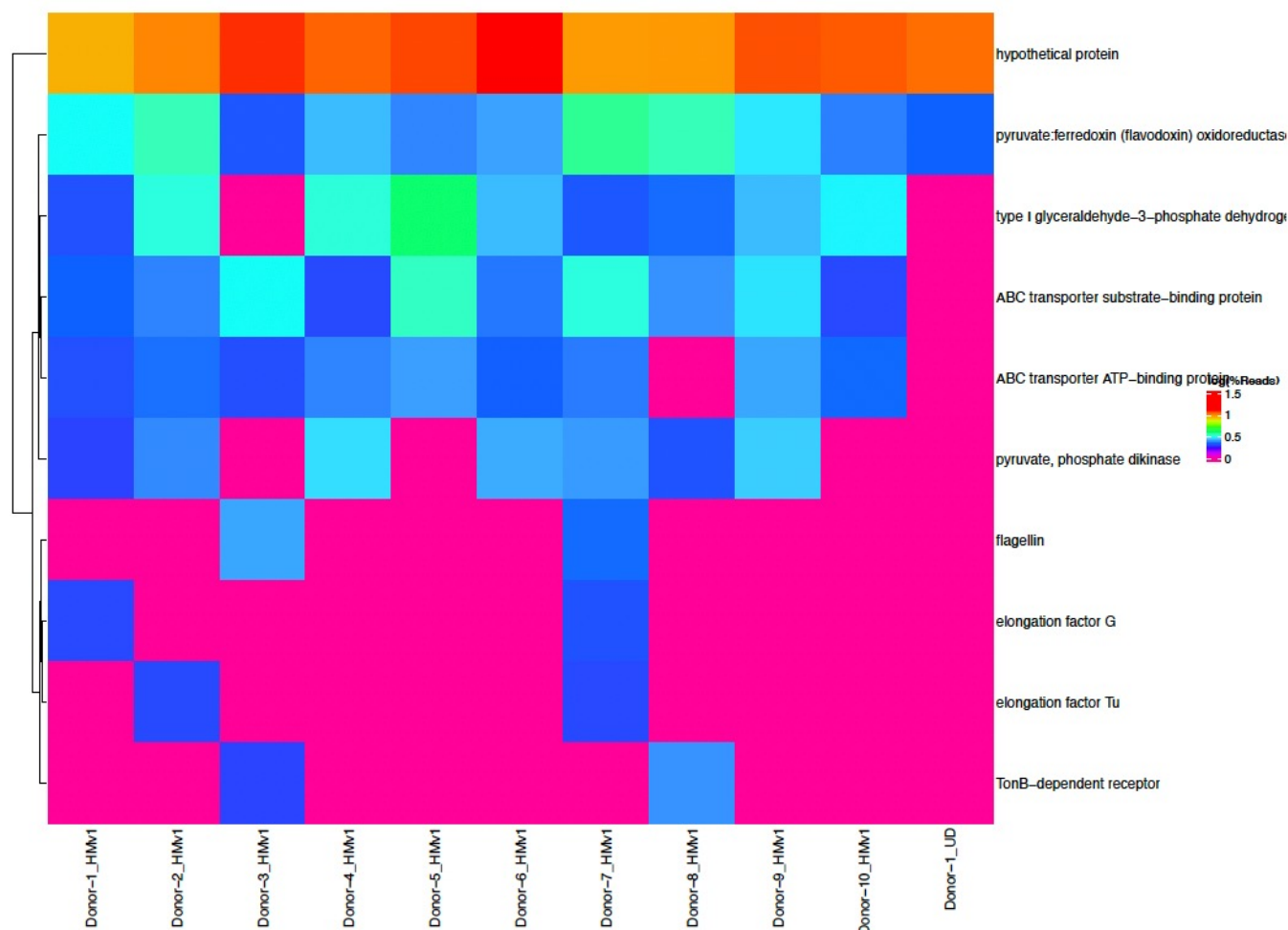


Figure S3

A



B

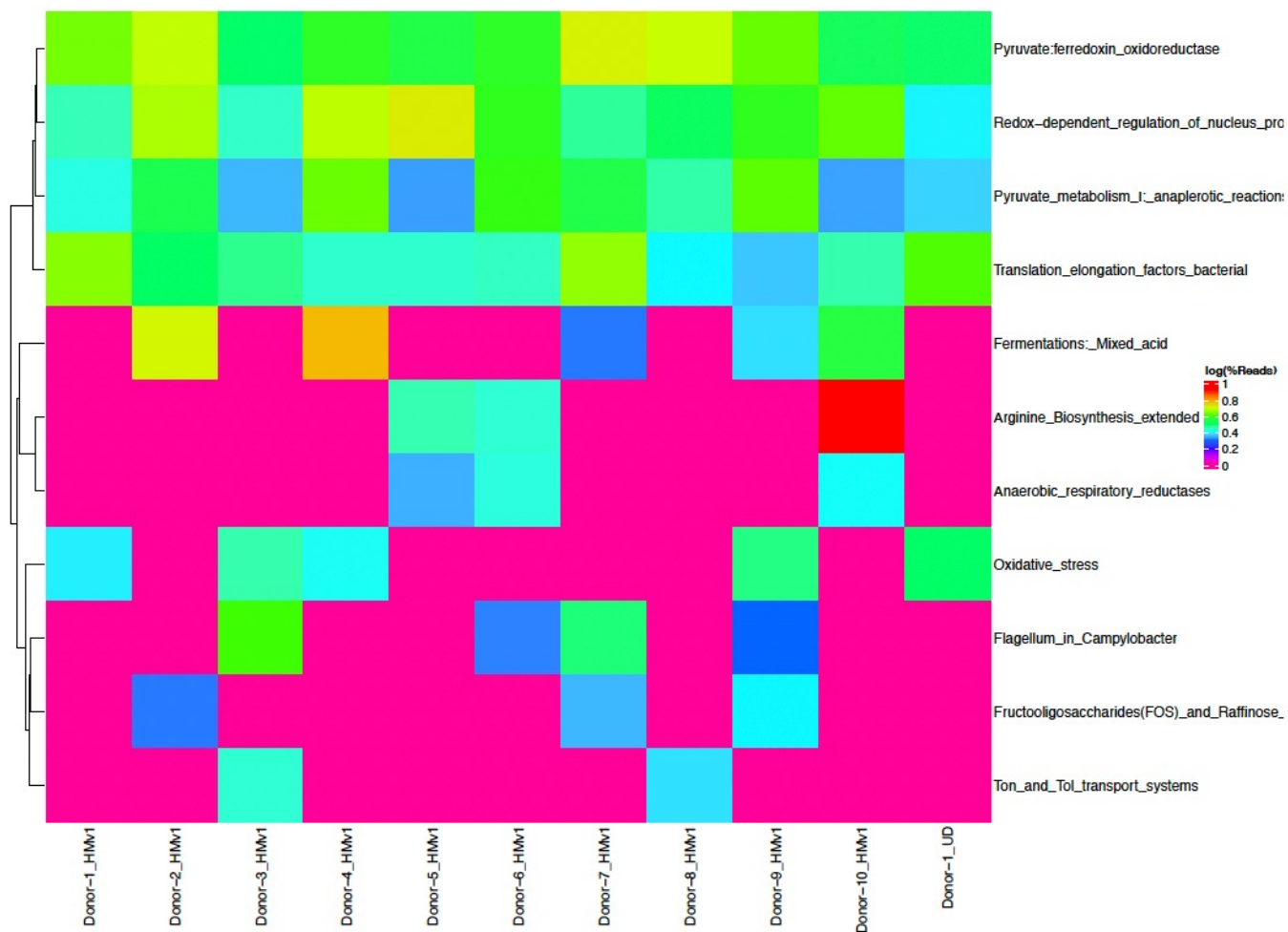
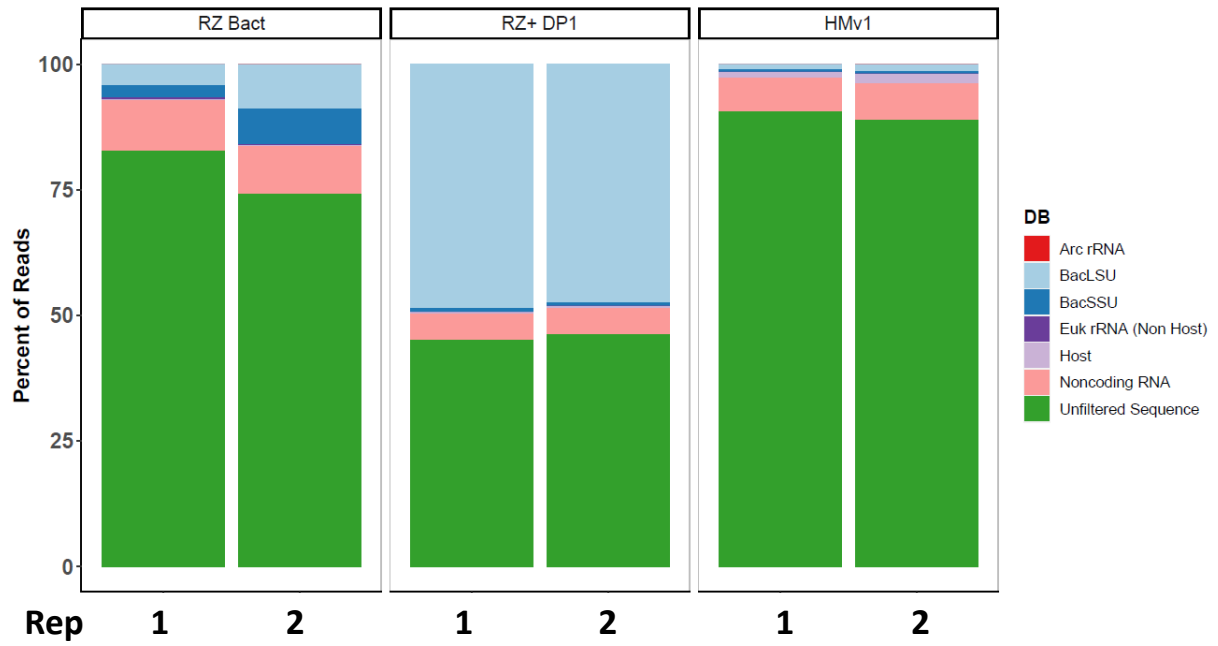


Figure S4

A. ATCC skin



B. ATCC gut

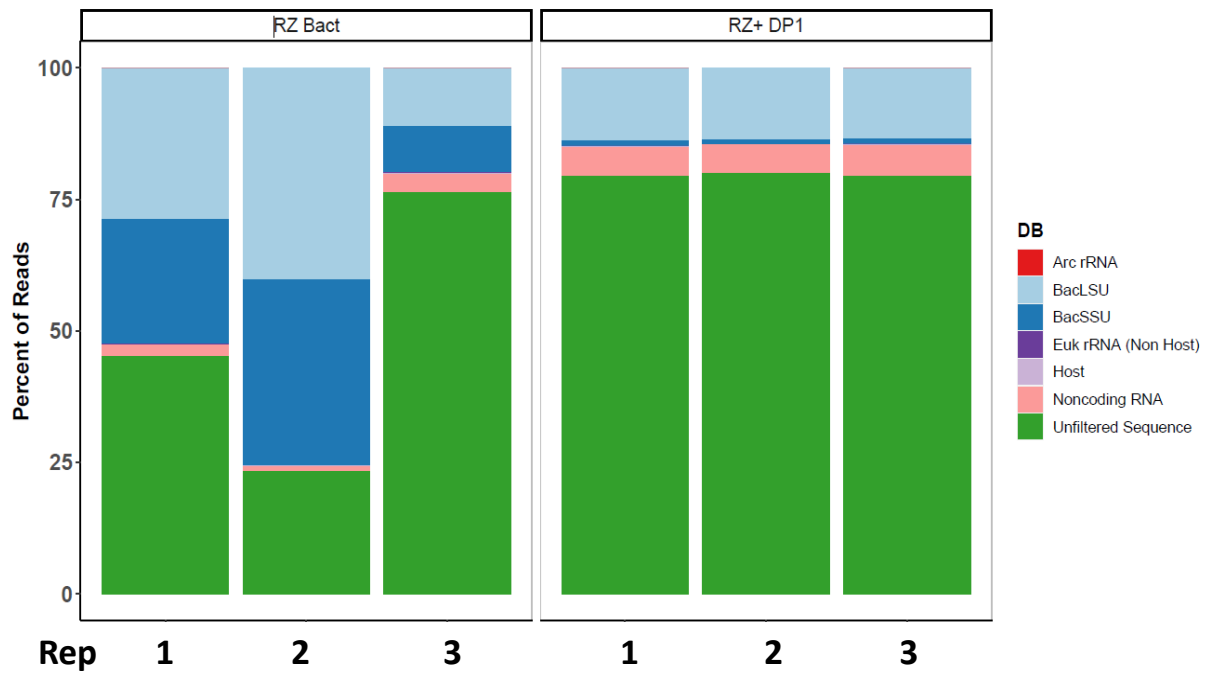


Figure S5

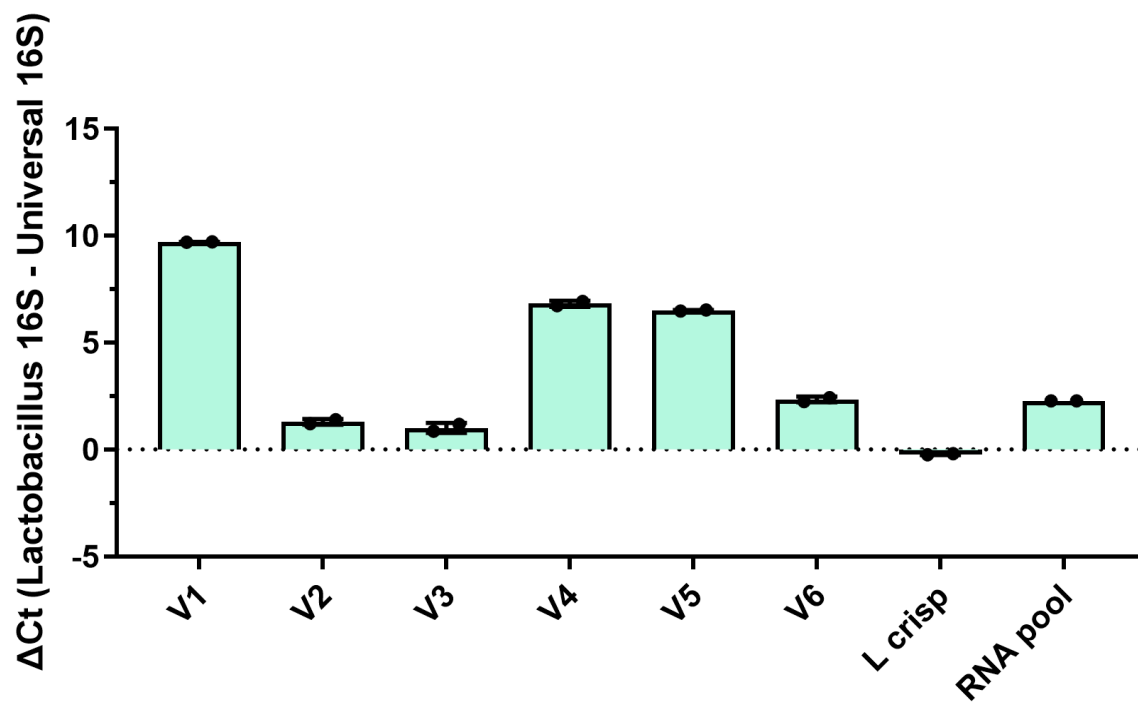
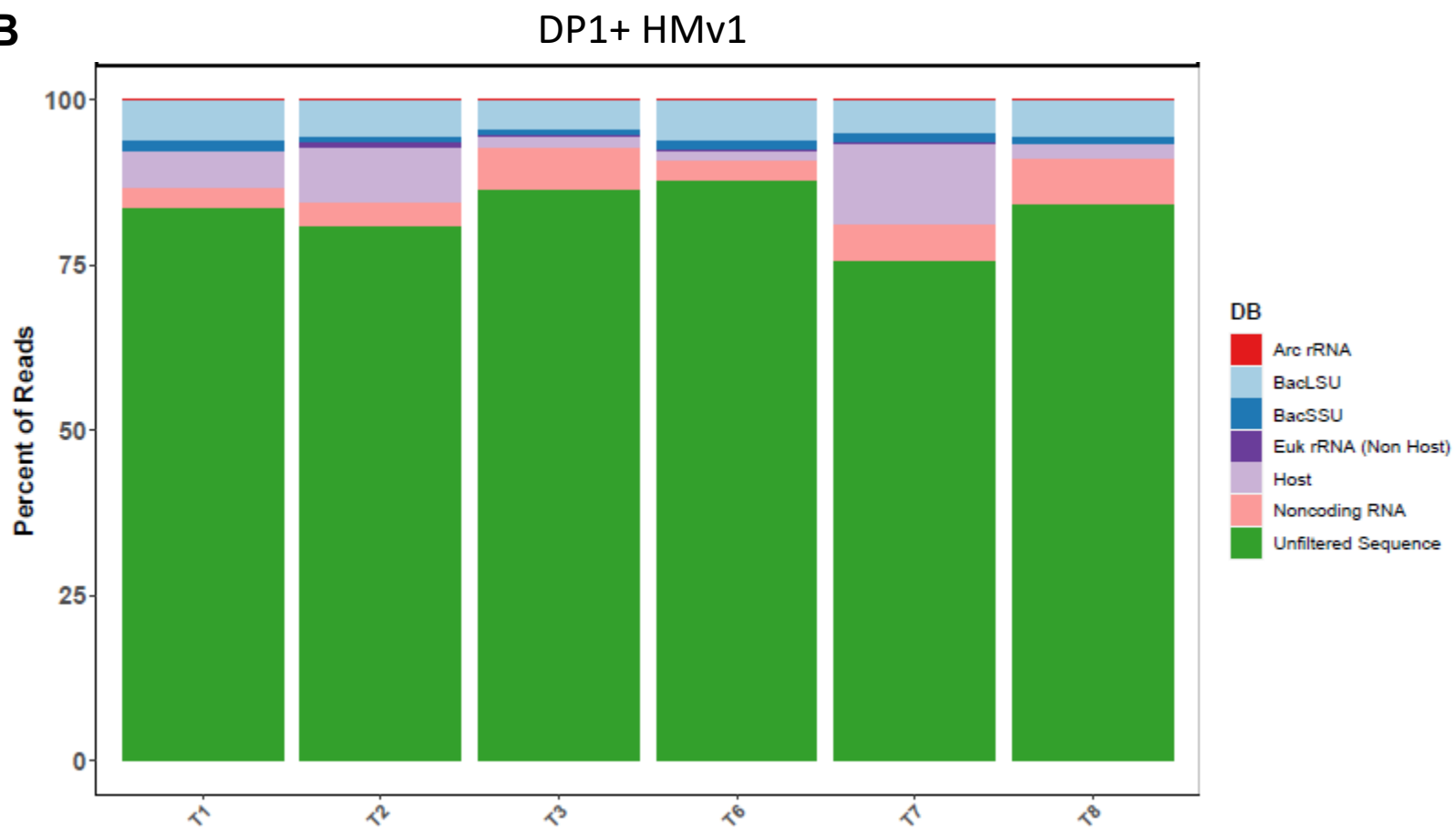
A**B**

Figure S6

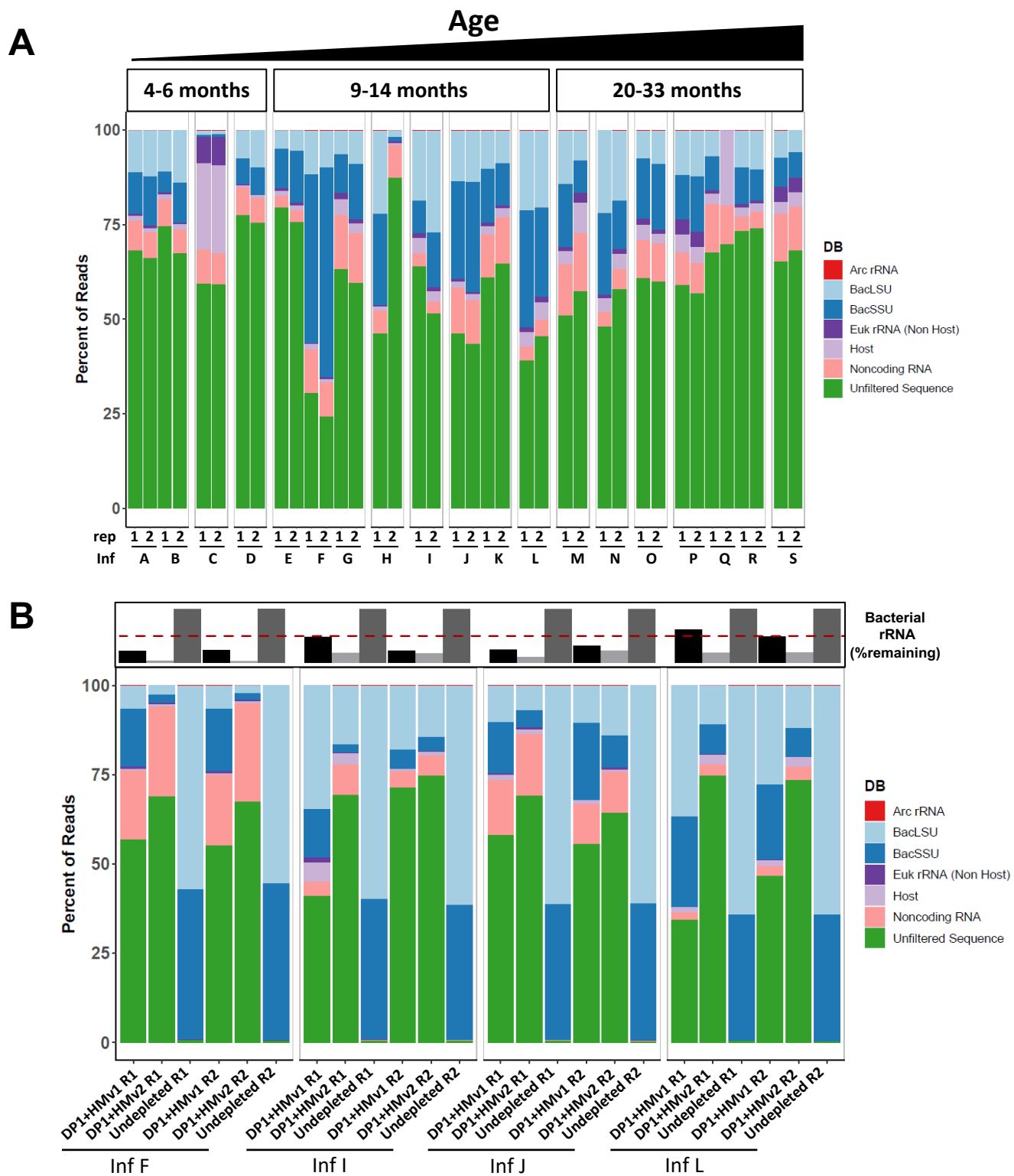


Figure S7

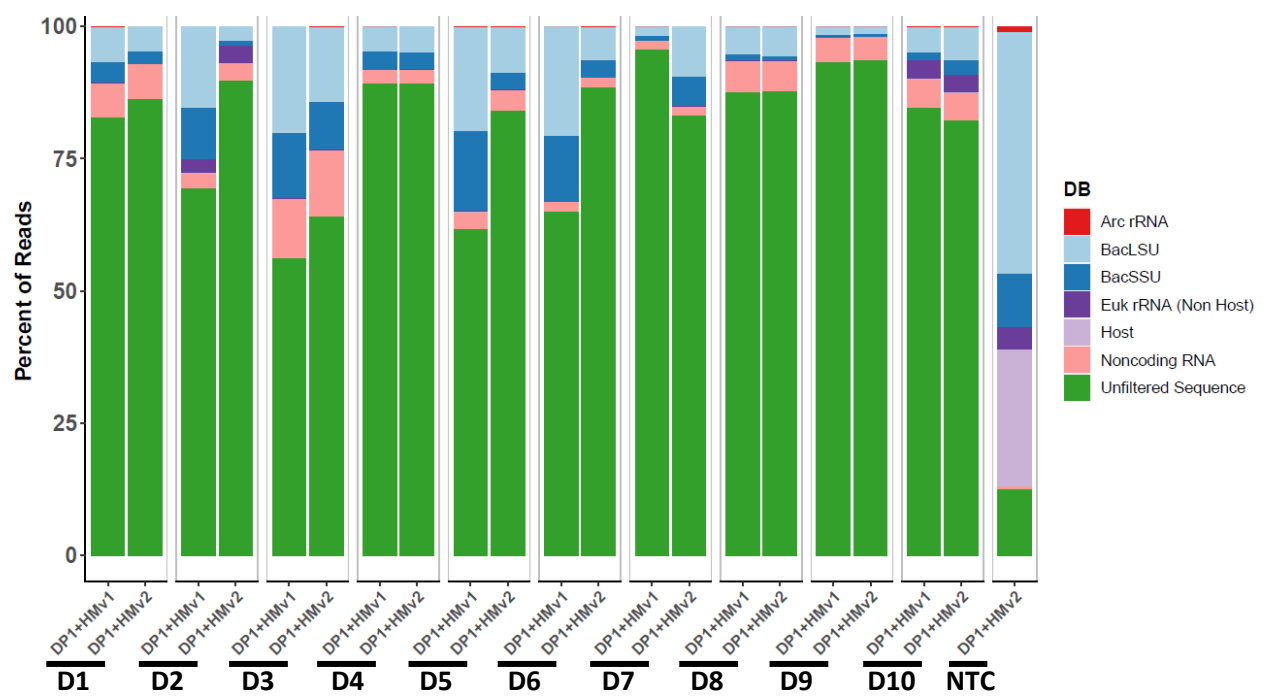


Figure S8

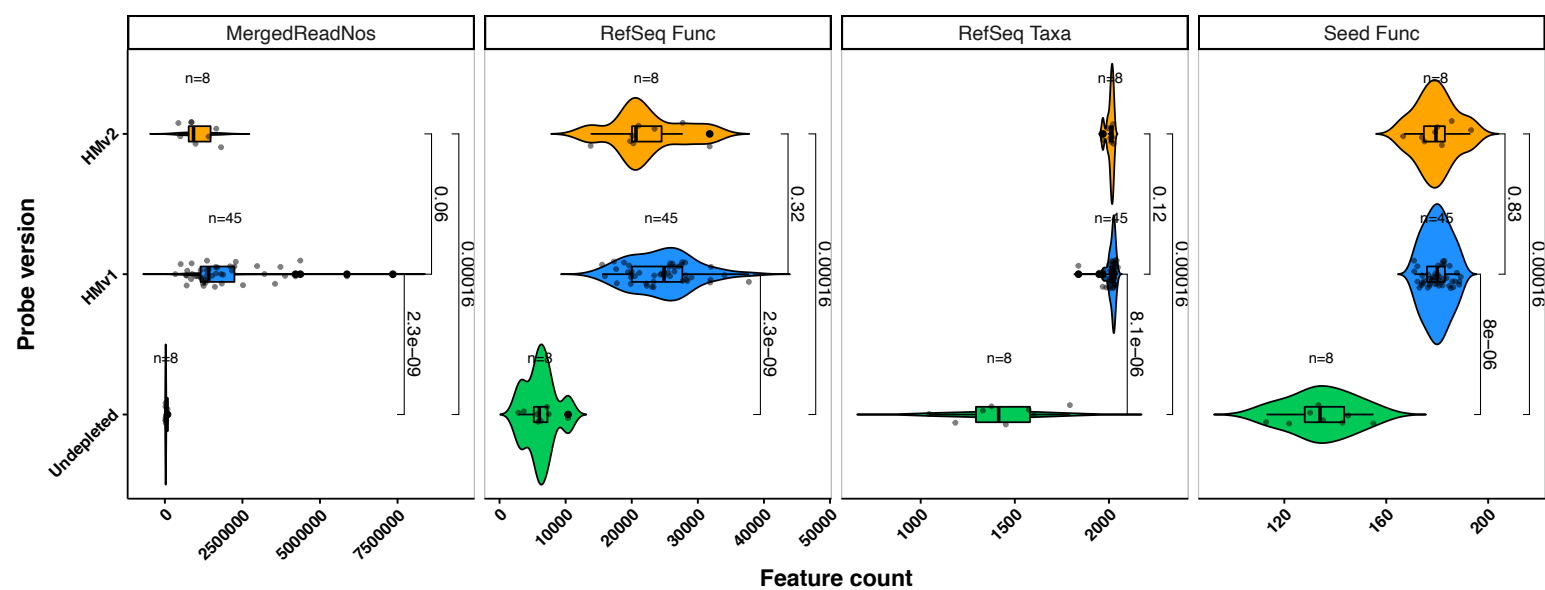


Figure S9

B

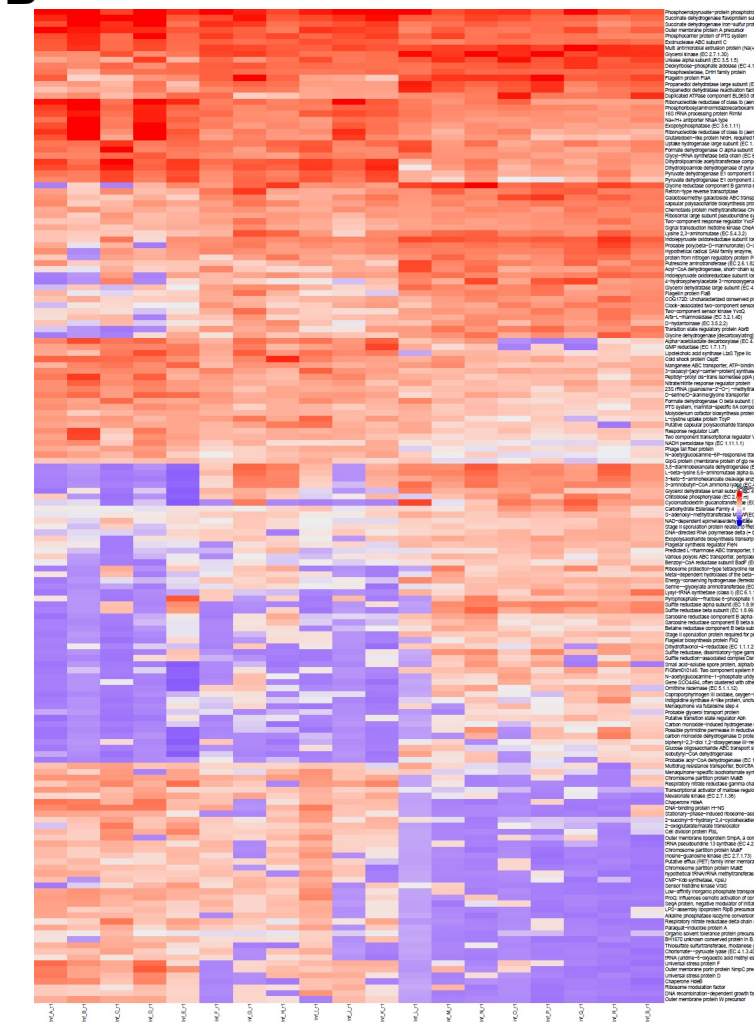


Figure S10

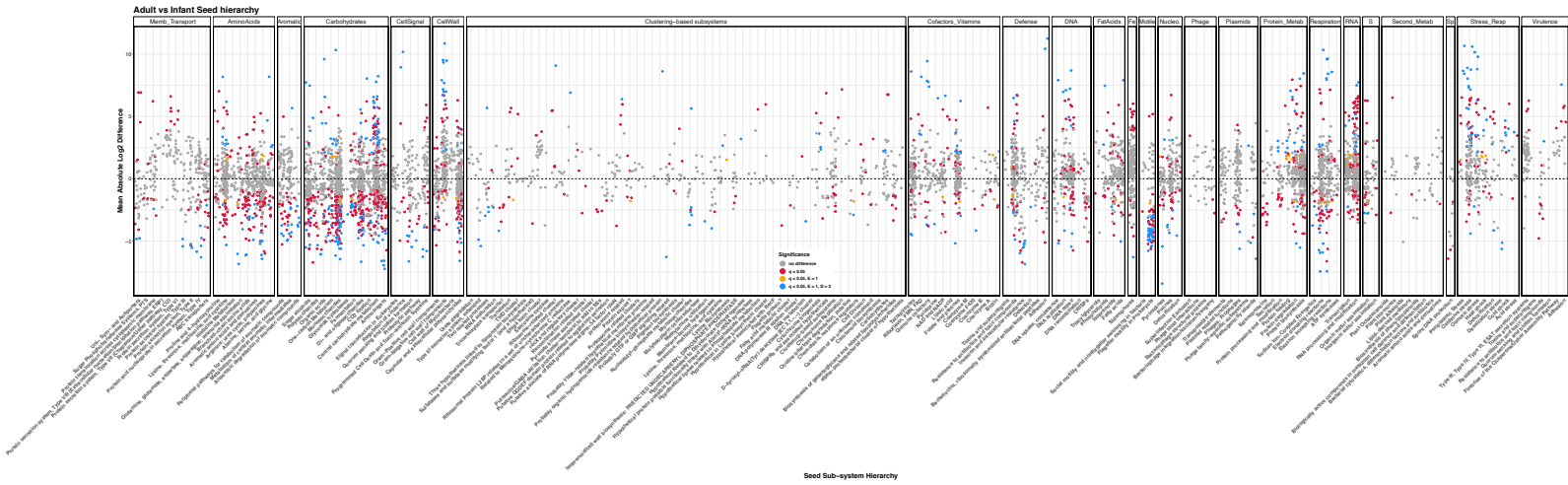
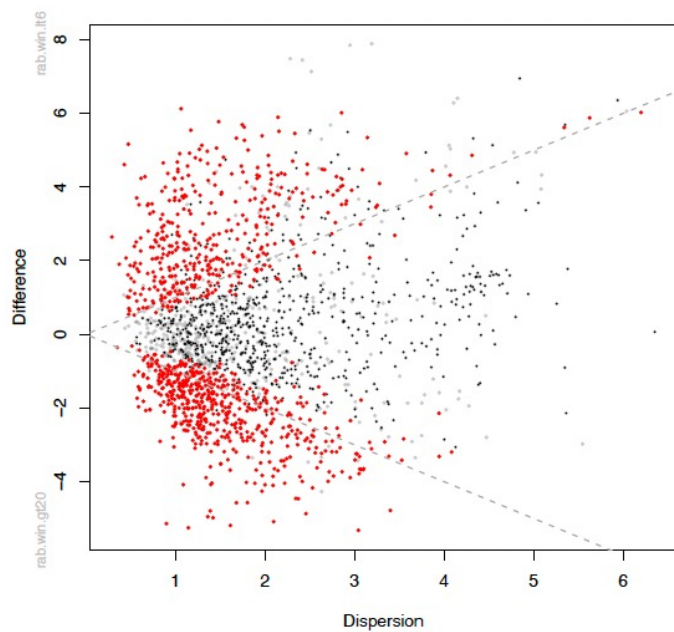
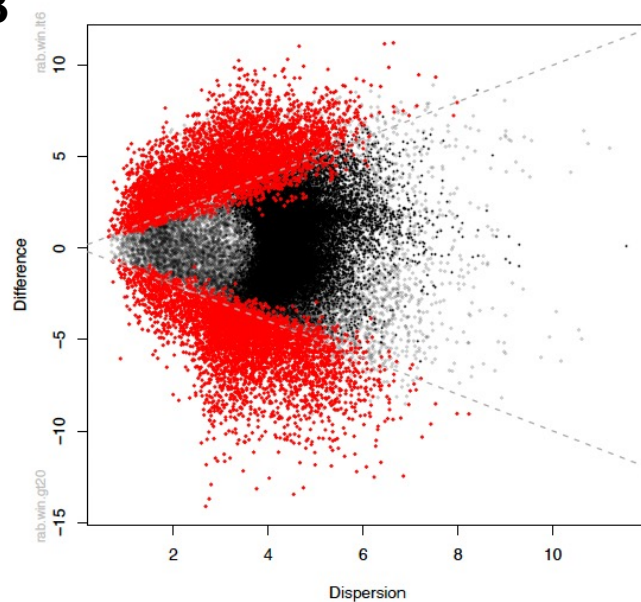
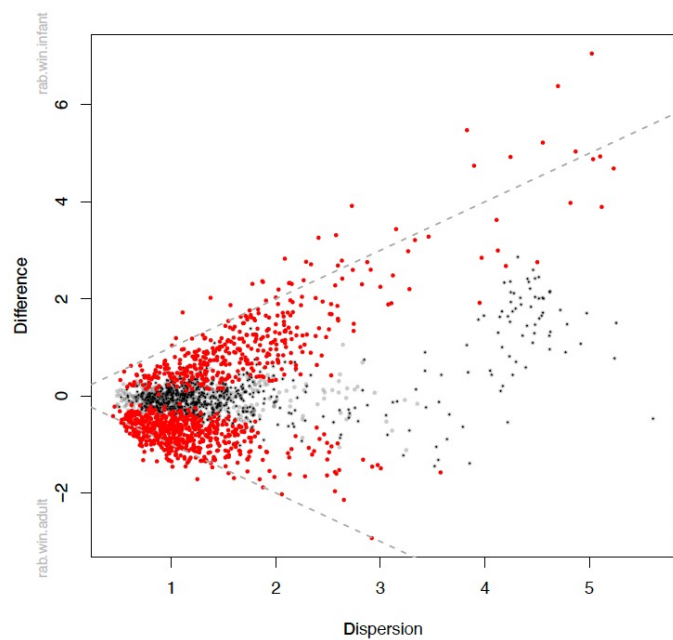
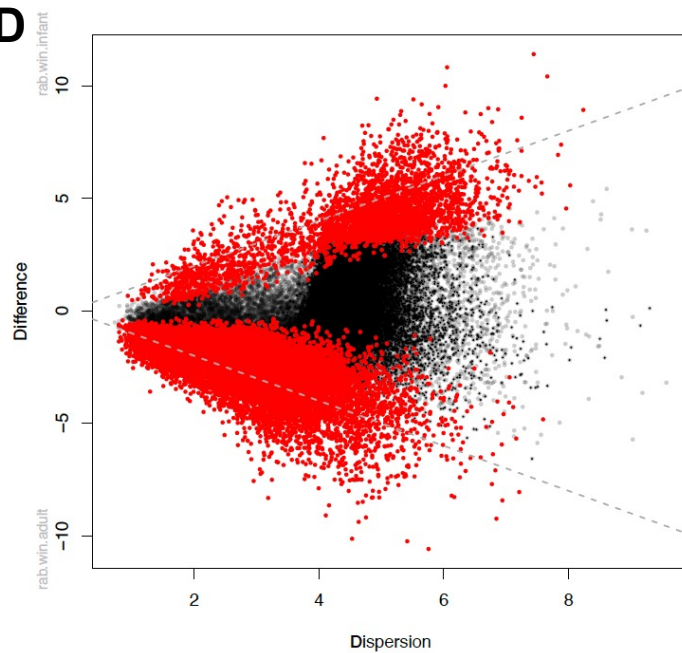
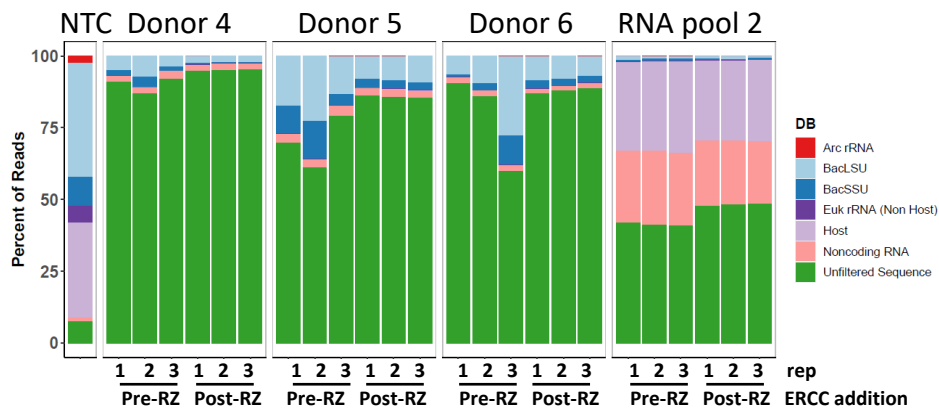


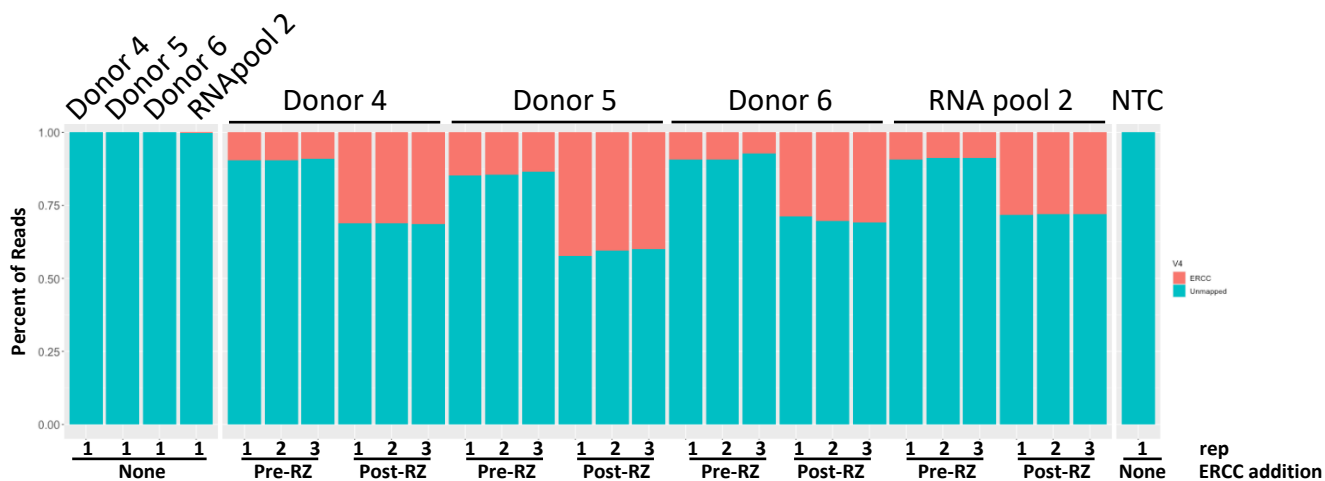
Figure S11

A**B****C****D**

A



B



C

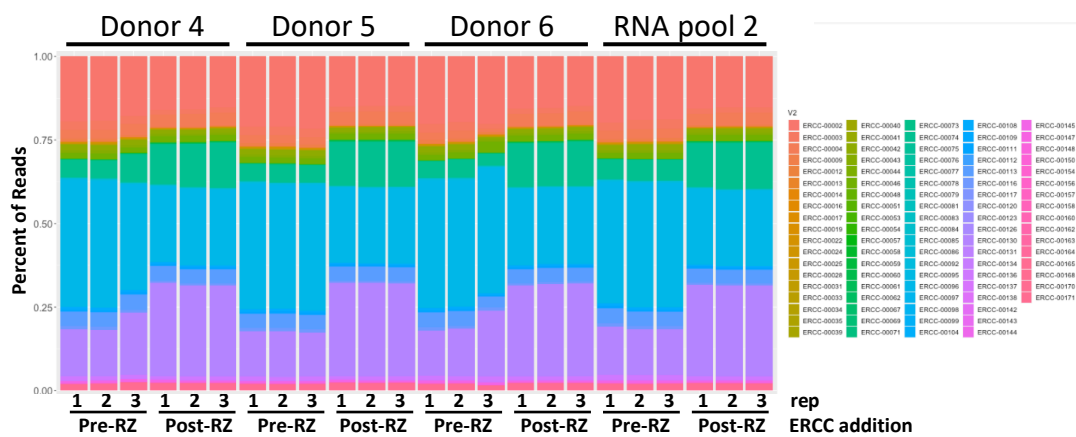


Figure S13

Variables Ordered and Colored by Spearman Correlation Coefficient, Replicate TPM Averaged

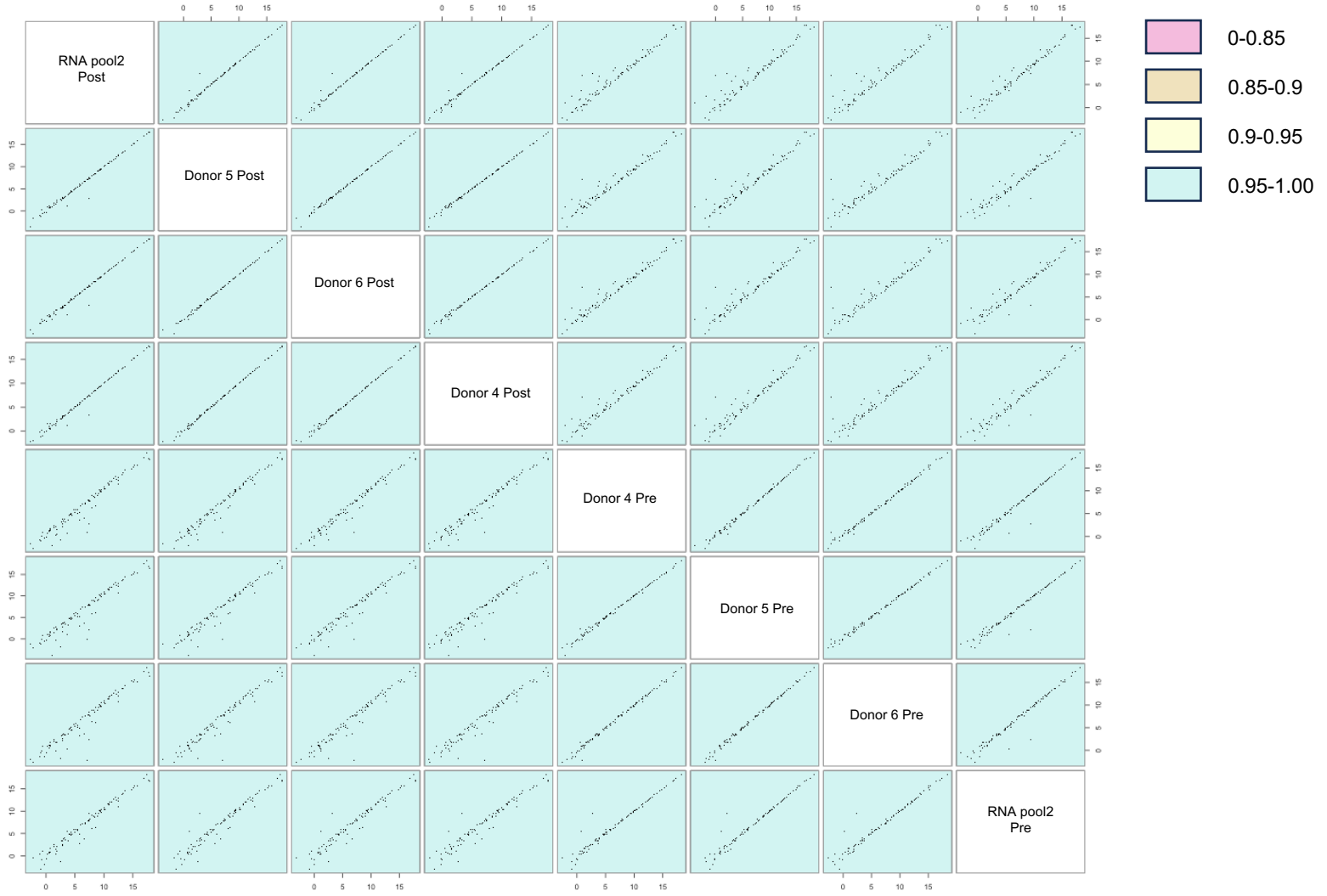


Figure S14